

Inner structure of vehicular ensembles and random matrix theory

Milan Krbálek and Tomáš Hobza

Faculty of Nuclear Sciences and Physical Engineering, Czech Technical
University in Prague, Prague, Czech Republic

E-mail: milan.krbalek@fjfi.cvut.cz

Abstract. We introduce a special class of random matrices (DUE) whose spectral statistics corresponds to statistics of microscopical quantities detected in vehicular flows. Comparing the level spacing distribution (for ordered eigenvalues in unfolded spectra of DUE matrices) with the time-clearance distribution extracted from various areas of the flux-density diagram (evaluated from original traffic data measured on Czech expressways with high occupancies) we demonstrate that the set of classical systems showing an universality associated with Random Matrix Ensembles can be extended by traffic systems.

PACS numbers: 05.40.-a, 89.40.-a, 45.70.Qj

Introduction

Connection between Random Matrix Ensembles and certain transport systems is not new (see e.g.[1] for a general overview). Indeed, in Ref.[2, 3] authors drawn the attention to the bus transportation system in Cuernavaca, Mexico, where a peculiar transfer of information inside the system led, surprisingly, to the universal configuration of gaps among buses. The rigorous study [4] confirmed that the detected link between Cuernavaca buses and Gaussian unitary ensemble (GUE) is not accidental. Authors of the research formulated a relevant microscopic model-scheme and analytically proved that the time headway distribution of buses conforms to the level spacing distribution for GUE. This surprising knowledge attracted an attention of scientists (e.g. [1]) since systems connected to Random Matrix Theory are universal (in a certain sense).

Until now, natural endeavor to interconnect Random Matrix Theory (RMT) with vehicular systems has not led to a success. However, a partial progress has been achieved in [5, 6, 7, 8, 9] where it is proved that microscopical arrangement of vehicles can be predicted by means of a certain one-dimensional gas inspired by the Dyson's gases that are well-known in RMT. Moreover, another attempts to describe (analytically) a microstructure of vehicular ensembles with help of statistical instruments (e.g. [10, 11, 12]) led us to believe that our effort may be successful. Therefore, the main goal of this work is to find (and analyze) a new class of random matrices whose spectral properties correspond to a micro-structure of real-road traffic samples.

This paper is organized as follows. In the first section we introduce and analyze the class of the so-called damped matrices (DUE). We are focused on the level density, the procedure of the unfolding, and the level spacing distribution. Connection to the theory of one-dimensional traffic gases is discussed in the second section. Section Three brings in a comparison between the level spacing distribution of DUE matrices and the time-clearance distributions extracted from expressway data samples. This is followed by corresponding discussion and conclusions in the final section. Subsidiary derivations are included in the mathematical Appendix.

1. DUE – Damped unitary ensembles

Being inspired by the work [13] (studying Calogero-Moser models with various potentials) we introduce the g -parameterized class $\text{DUE}_g(N)$ whose $N \times N$ matrices $H_g = (h_{kj}(g))_{k,j=1}^N$ fulfil the following axioms:

- (i) the parameter $g \geq 0$ is fixed;
- (ii) elements $h_{kk}(g)$ are chosen as independent Gaussian variables with zero mean and unit variance, i.e. $h_{11}(g), h_{22}(g), \dots, h_{NN}(g) \sim \text{N}(0, 1)$ are i.i.d.;
- (iii) if $k \neq j$ then

$$h_{kj}(g) = \frac{2\pi i g}{N \sinh[(2\pi(k-j)/N)],}$$

i.e. the off-diagonal elements are deterministic.

Owing to the definition the off-diagonal elements are purely imaginary complex numbers whose absolute value become smaller with distance from the diagonal. Therefore we refer such a class to as the Damped Unitary Ensembles. In the context of the article [13] the matrices in $\text{DUE}_g(N)$ represent simplified versions of the $N \times N$ Lax matrices derived for hyperbolic Calogero-Moser model (see [13] for details).

1.1. Level density for DUE matrices

Let $\sigma(H_g)$ be the spectrum of H_g , i.e. $\sigma(H_g) = \{x_{(k)} \in \mathbf{C} : k = 1, 2, \dots, N\}$ is a set of ordered eigenvalues $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$. Since $H_g = H_g^\sharp$, where H_g^\sharp is Hermitian conjugated matrix, one finds that $\sigma(H_g) \subset \mathbf{R}$ and therefore the above-referred ordering is correctly defined. Now we can introduce the probability density $P_k(x_{(k)})$ for k th eigenvalue $x_{(k)}$ and the level density

$$\omega_N(x) = \frac{1}{N} \sum_{k=1}^N P_k(x). \quad (1)$$

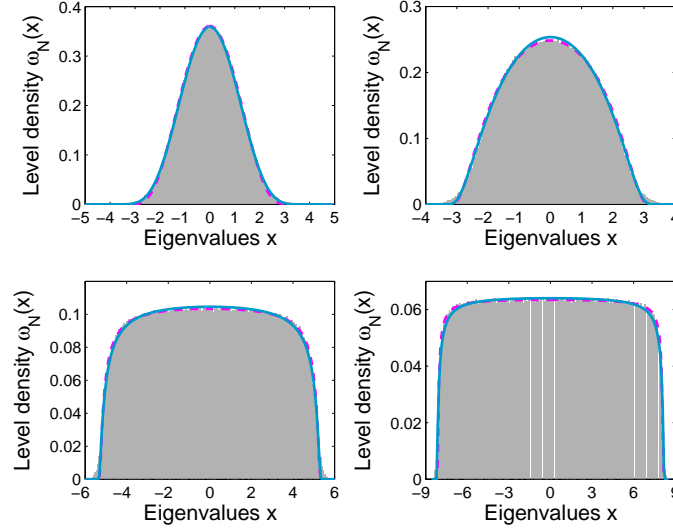


Figure 1. Graphs of the level density $\omega_N(x)$. Histograms have been determined for five thousand realizations of DUE(256) matrices with parameters $g = 0.2$ (a northwest sub-figure), $g = 0.5$ (a northeast sub-figure), $g = 1.5$ (a southwest sub-figure), and $g = 2.5$ (a southeast sub-figure). Solid curve (blue)/dashed curve (magenta) represent the approximation (2) for values ε, ϑ estimated with help of CSC/KE methods, respectively.

Unfortunately, the famous Semi-Circle Law (derived in [14] for the level density of classical random matrix ensembles) is not applicable for DUE matrices. Instead of a circular shape the level density of damped matrices conforms to a probability density taken from two-parametric family

$$q_{\vartheta, \varepsilon}(x) = \frac{\zeta(\vartheta)}{\varepsilon} \begin{cases} \exp[-\frac{\vartheta^2 \varepsilon^2}{\varepsilon^2 - x^2}] & \dots \quad |x| < \varepsilon, \\ 0 & \dots \quad |x| \geq \varepsilon, \end{cases} \quad (2)$$

where $\zeta^{-1}(\vartheta) = \int_{-1}^1 \exp[-\frac{\vartheta^2}{1-x^2}] dx$ ensures the proper normalization. Indeed, numerical tests show (as is illustrated in figure 1) that the level density for large N is very accurately approximated by the function $q_{\vartheta, \varepsilon}(x)$ for estimated values $\hat{\varepsilon} = \hat{\varepsilon}(g, N)$, $\hat{\vartheta} = \hat{\vartheta}(g, N)$ of the so-called critical eigenvalue ε and curvature ϑ . Both the presented estimation-procedures are based on principles of MDE (minimum distance estimation). The first of them (CSE – chi-square estimator) minimizes the statistical distance

$$\chi(\vartheta, \varepsilon) = \sqrt{\int_{-\infty}^{+\infty} |h(x) - q_{\vartheta, \varepsilon}(x)|^2 dx}, \quad (3)$$

where $h(x)$ is an empirical histogram. For potential reconstructive purposes we are specifying that the set S consisting of 1 280 000 eigenvalues (taken from 5 000 realizations of 256×256 matrices) generates the domain $\text{Dom}(h) = [\min(S), \max(S)]$ of the empirical histogram. In our numerical tests, this domain $\text{Dom}(h)$ has been divided into 200 equidistant sub-intervals.

The second statistical test is based on the standard Kolmogorov estimator (KE) minimizing the supremum of the absolute difference between the estimated distribution function and the empirical distribution function

$$H_N(x) = \frac{1}{N} \sum_{k=1}^N I(x_{(k)} \leq x). \quad (4)$$

Here, $I(x)$ stands for the indicator function. The optimal values of estimated parameters are compared graphically in figure 2. Marked discrepancies near the origin can be attributed to the following facts. If $g = 0$ then one can trivially express the level density as $e^{-\frac{x^2}{2}}/\sqrt{2\pi}$ and our estimation procedure tries, in fact, to approximate the normal distribution by a function $q_{\vartheta, \varepsilon}(x)$. Moreover,

$$\forall \vartheta > 0: \quad \lim_{\varepsilon \rightarrow 0_+} q_{\vartheta, \varepsilon}(x) = \lim_{x \rightarrow 0_+} \frac{1}{\sqrt{2\pi x}} e^{-\frac{x^2}{2x}} = \delta(x),$$

where $\delta(x) \in \mathcal{D}'$ is the Dirac function (see Appendix 5.1). Since both these distributions tend (for small values of parameters) to the same generalized limit it is difficult to distinguish them numerically. It means that estimations used (applied for small values of g) are highly sensitive on a type of estimators, quality of analyzed data and so on. On the other hand, nuances in parameters estimated by different methods cause only imperceptible deviations in curves of the level density (see solid and dashed lines in the northwest sub-figure in figure 1)).

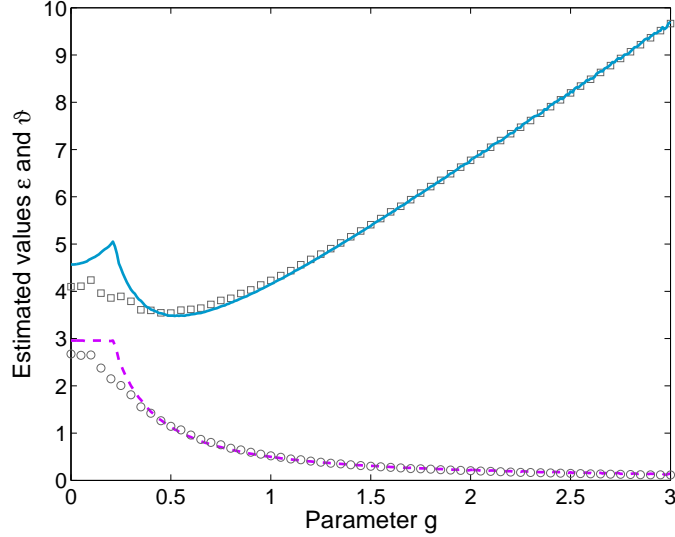


Figure 2. Estimates of the critical eigenvalue ε and curvature ϑ in (2). A curve and dashed line correspond to $\hat{\varepsilon}$ and $\hat{\vartheta}$ obtained by CSE method. Squares and circles display the estimates $\hat{\varepsilon}$ and $\hat{\vartheta}$ calculated by means of KE.

1.2. Procedure of the unfolding

Before statistical analysis of random-matrix spectra the standard methodology of Random Matrix Theory requires a transition to the so-called *unfolded spectrum*. The main benefit of that transition is an unified view into a structure of heterogeneous systems. To be specific, the procedure of the unfolding converts eigenvalue-spectrum $\sigma(H_g)$ with an arbitrary level density $\omega_N(x)$ (depending on specific properties of a system) to the unfolded spectrum $\tilde{\sigma}(H_g) = \{\tilde{x}_{(k)} \in \mathbf{C} : k = 1, 2, \dots, N\}$ in which unfolded eigenvalues are uniformly distributed in $(0, N)$. Thus, the level density of the unfolded spectrum is associated with the uniform distribution $U(0, N)$. Mathematically, the unfolding is a procedure consisting of two operations: an individual mapping $x_{(k)} \mapsto \bar{x}_{(k)}$ defined by

$$\bar{x}_{(k)} := \int_{-\infty}^{x_{(k)}} \omega_N(y) dy, \quad (5)$$

and a global scaling

$$\tilde{x}_{(k)} := \frac{\bar{x}_{(k)}(N-1)}{\sum_{k=1}^{N-1} (\bar{x}_{(k+1)} - \bar{x}_{(k)})}. \quad (6)$$

According to the theorem on probability integral transformation (see [15]) it holds $\bar{x} \sim U(0, 1)$, which results (for large N) in the fact $\tilde{x} \sim U(0, N)$. Moreover, the average gap between two succeeding unfolded eigenvalues is equal to one.

Specifications: All the previous considerations require a large value of N . Thus, in our numerical realizations we consistently use $N = 256$.

1.3. Level spacing distribution

In this subsection we aim to deduce an analytical formula predicting the probability density for gaps $s_k := \tilde{x}_{(k+1)} - \tilde{x}_{(k)}$ between the ordered eigenvalues $\tilde{x}_{(1)} \leq \tilde{x}_{(2)} \leq \dots \leq \tilde{x}_{(N)}$ in the unfolded spectrum

$\tilde{\sigma}(H_g)$. Here, such a function is referred as the *level spacing distribution (LS-distribution)* and denoted by $\wp(s)$.

The most trivial variant of ensembles investigated is the class $\text{DUE}_0(N)$ of totally damped matrices. The matrices $H \in \text{DUE}_0(N)$ are diagonal and therefore the corresponding level density reads $\omega_N(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for all N . Thus, the first stage of the unfolding-procedure

$$\bar{x} := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy = \frac{1}{2} \left(1 + \text{Erf}\left[\frac{x}{\sqrt{2}}\right] \right) \quad (7)$$

produces i.i.d.-numbers from the uniform distribution $U(0,1)$. Hence, one can easily verify that level spacing distribution in $\text{DUE}_0(N)$ can be expressed as

$$\wp(s) = \Theta(s) e^{-s}. \quad (8)$$

In the article [13] authors derived that the joint probability density for eigenvalues in the hyperbolic Calogero-Moser model corresponds to the law

$$P(x_1, x_2, \dots, x_N) \propto \exp \left[-\Omega \sum_{k=1}^N x_k^2 \right] \prod_{k=1}^N \mathcal{K}_0 \left(\Psi \prod_{j \neq k} \left| 1 + \frac{4\pi i g}{N(x_j - x_k)} \right| \right), \quad (9)$$

where Ω, Ψ are constants and $\mathcal{K}_m(x)$ stands for the Macdonald's function of the order m – solution of the modified Bessel's differential equation of the second kind $x^2 y'' + xy' - (x^2 + m^2)y = 0$. Having regard to the fact that $\text{DUE}_g(N)$ matrices can be understood as numerical simplifications of Lax matrices in the hyperbolic Calogero-Moser model (as is discussed in [13]), the function (9) can be considered as a suitable approximation for the joint probability density in the ensemble $\text{DUE}_g(N)$. Although the expression (9) is difficult to handle, it allows to obtain a Wigner-type surmise for the level spacing distribution in $\text{DUE}_g(N)$. Considering $N = 2$ (similarly to the original Wigner idea) we have

$$P(x_1, x_2) \propto e^{-\Omega(x_1^2 + x_2^2)} \mathcal{K}_0^2 \left(\Psi \sqrt{1 + \frac{4\pi^2 g^2}{(x_2 - x_1)^2}} \right). \quad (10)$$

From the general relation $\ell(r) = \Theta(r) \left(\int_{\mathbf{R}} P(r+x, x) dx + \int_{\mathbf{R}} P(x, r+x) dx \right)$ we can obtain the probability density

$$\ell(r) \propto \Theta(r) \mathcal{K}_0^2 \left(\Psi \sqrt{1 + \frac{4\pi^2 g^2}{r^2}} \right) e^{-\Omega r^2/2} \quad (11)$$

for a distance between non-unfolded eigenvalues in $\text{DUE}_g(2)$. Application of the approximation $\mathcal{K}_0(z) \approx \sqrt{\frac{\pi}{2z}} e^{-z} + \mathcal{O}(1/z)$ is resulting in the formula

$$\ell(r) \propto r e^{-\frac{4\pi\Psi g}{r}}, \quad (12)$$

describing a behavior of the distribution studied at small distances r . An immediate consequence of this result is evidence of repulsions between eigenvalues that are rapidly stronger than those detected in the standard matrix ensembles (i.e. **GOE**, **GUE**).

Combining the two asymptotic predictions (8) and (12) we suggest (analogously to [13]) the level spacing distribution in the form

$$\wp(s) = A s^\alpha e^{-\beta/s - Ds}, \quad (13)$$

where (according to Appendix 5.2)

$$D = \beta + \alpha + \frac{3 - e^{-\sqrt{\beta}}}{2}, \quad (14)$$

$$A^{-1} = 2 \left(\frac{\beta}{D} \right)^{\frac{\alpha+1}{2}} \mathcal{K}_{\alpha+1}(2\sqrt{\beta D}). \quad (15)$$

On contrary to the pioneering work [13] our paper intends to investigate the level spacing distribution of **DUE**-matrices in more details. As the curve of the level density near the values $x = \pm \varepsilon$ is extremely steep (see, for example, the southeast sub-figure 1), the numerical representation of the unfolding (using an approximate formula (2) for the level density) is unstable in those regions. For this reason we analyze a truncated spectrum $\tilde{\sigma}_m(H_g) = \{\tilde{x}_{(k)} : k = m+1, 2, \dots, N-m\}$ whose elements are re-scaled so that the mean spacing is equal to one. Then the set $\{s_{m+1}, s_{m+2}, \dots, s_{N-m-1}\}$ of empirical spacings is subjected

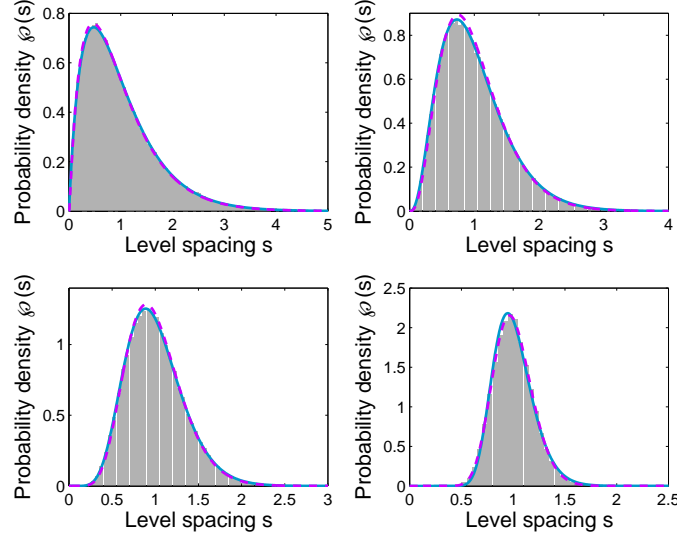


Figure 3. Level spacing distribution. The gray histograms correspond to empirical LS-distributions of $\text{DUE}_g(256)$ obtained for one thousand matrix-realizations and for $g = 0.2$ (a northwest sub-figure), $g = 0.5$ (a northeast sub-figure), $g = 1.0$ (a southwest sub-figure), and $g = 2.0$ (a southeast sub-figure). The continuous/dashed curves display the prognosis (13), in which the parameters have been estimated by means of MDE/MLE estimators, respectively. For all variants the value $m = 50$ has been used.

to standard procedures estimating optimal values α, β in the functional premise (13). Here we present two standard estimators: MLE – maximum likelihood estimation and MDE – minimum distance estimation. For these purpose we introduce the log-likelihood function

$$\mathcal{L}(\alpha, \beta) = (N-2m-1) \ln(A) + \alpha \sum_{k=m+1}^{N-m-1} \ln(s_k) - \beta \sum_{k=m+1}^{N-m-1} \frac{1}{s_k} - D \sum_{k=m+1}^{N-m-1} s_k \quad (16)$$

and the weighted statistical distance

$$\chi(\alpha, \beta) := d[h, \varphi] = \sqrt{\int_0^{+\infty} \frac{\pi s}{2} e^{-\frac{\pi s^2}{4}} |h(s) - \varphi(s; (\alpha, \beta))|^2 ds}, \quad (17)$$

where $h(s)$ is an empirical histogram for spacings. We remark that the weight factor $\varpi(s) := \frac{\pi s}{2} e^{-\frac{\pi s^2}{4}}$ has been chosen to eliminate the influence of long spacings and increase the influence of spacings that are close to the mean value. In addition, $\arg\max_{s \geq 0} \varpi(s) = 1$ and $d[0, 1] = 1$, which can be understood (if d is now understood as a general functional metric) as a benchmark. As graphically illustrated in figure 3 LS-distributions of DUE -matrices is in full compliance with the analytical prediction (13). The optimal values of estimated parameters, i.e. $(\hat{\alpha}, \hat{\beta})_{\text{MLE}} = \arg\max \mathcal{L}(\alpha, \beta)$, $(\hat{\alpha}, \hat{\beta})_{\text{MDE}} = \arg\min \chi(\alpha, \beta)$ are then compared graphically in figure 4.

2. Local thermodynamical traffic model

As one of the main goals of this article is to find a connection between the above-referred matrix ensembles and freeway traffic samples, it is beneficial to show that the functional premise (13) can be understood as a steady-state distribution of particle headways in the thermal-like socio-physical traffic model with short-ranged repulsions between particles (originally introduced in [16]). Indeed, if considering circular ensemble of particles whose mutual distances (between subsequent particles) are denoted by s_1, s_2, \dots, s_n and if understanding such an ensemble as thermodynamical (with a thermal parameter $\nu \geq 0$) we are able to derive (following the approaches presented in [16]) that the headway distribution in the ensemble with a repulsive potential energy

$$U_\mu(s_1, s_2, \dots, s_n) = \sum_{k=1}^n \left(\frac{1}{s_k} - \mu \ln(s_k) \right), \quad (\mu \geq 0) \quad (18)$$

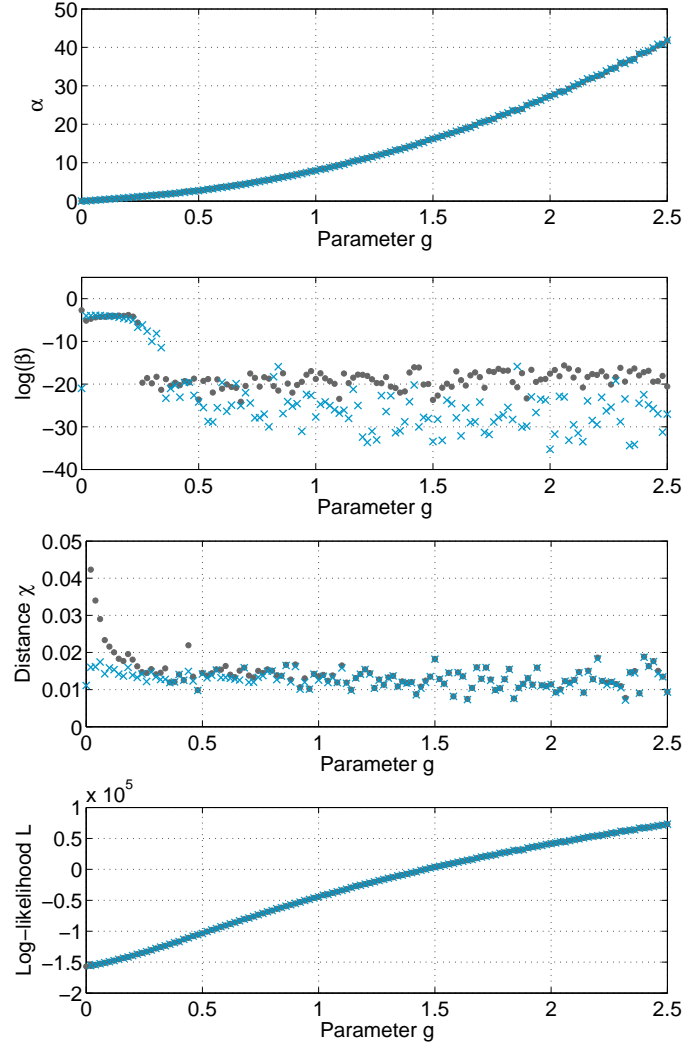


Figure 4. Estimated values of parameters in the functional premise (13). We visualize the parameters α, β estimated by the maximum likelihood estimation (bullets) and minimum distance estimation (crosses), respectively. Values of the statistical distance $\chi(\alpha, \beta)$ between the probability density (13) specified for estimated parameters and empirical histograms are plotted in the third sub-figure. Values of the log-likelihood $\mathcal{L}(\alpha, \beta)$ corresponding to the probability density (13) specified for estimated parameters are plotted in bottom sub-figure.

takes the form $\varphi(s) = As^{\mu\nu}e^{-\frac{\nu}{s}}e^{-Ds}$. Such a probability density fully corresponds to the generalized inverse Gaussian distribution (13) and moreover, it fulfills all the substantial criteria for acceptability of analytical clearance distributions presented in the articles [8, 9]. Indeed, for all values of $\mu \geq 0$ and all $\nu > 0$ the function $\varphi(s)$ generates the origin plateau, which means that it obeys the condition $q \in \mathbf{R}^+ \Rightarrow \lim_{s \rightarrow 0^+} s^{-q}\varphi(s) = 0$. Furthermore, the function $\varphi(s)$ is belonging to the class of balanced distributions (see [9]), which (together with the mathematical criteria **(T1–T5)** discussed in [9]) completes all theoretical requirements for vehicular headway distributions.

Since the presented thermodynamical traffic model represents a mathematically well-defined (and analytically soluble) alternative to the more realistic traffic model (see the discussion on the pages 6,7 in [8]) we have now a theoretical permission to use the above-referred statistics for a mathematical description of headway distributions in realistic vehicular streams.

3. Headway statistics in various regions of flux-density map

In this section we try to answer the fundamental question of this work: "Is there any link between fluid vehicular locations and eigenvalues in random matrices (similar to the link between Mexico buses and GUE-matrices presented in [2, 3, 4, 1])?"

3.1. On the origin of traffic data used

Vehicle-by-vehicle data analyzed in the paper have been recorded at the Expressway R1 (also called the Prague Ring) in Prague, the Czech Republic. For our measurements (conducted in cooperation with The Road and Motorway Directorate of the Czech Republic) we have chosen expressway-segments where traffic is strongly saturated and where therefore one can regularly detect synchronized flows. Having regard to intentions of this article we eliminate all free-flow states (where car-car interactions are negligible and therefore traffic behaves almost like poissonian system). For distinguishing congested-traffic states from free-flow states (see [18, 19]) we have applied the clustering algorithm K-means (see [17]). Here the chosen alternative of the method is based on minimization with respect to velocity of vehicles.

3.2. The traffic alternative for the spectral unfolding

Data record (adapted into a form suitable for further processing) has been divided into sub-samples $T_j = \{t_{j1}, t_{j2}, \dots, t_{jm}\}$ containing m consecutive time clearances (netto time intervals between succeeding cars passing a traffic detector located at the chosen lane), where m is the sampling size. For each sub-sample there were calculated the local flux J_j and local density ϱ_j . The method used for those calculations is identical to the standard method described mathematically in [8] (page 4). Again, we emphasize that all sub-samples belong to a regime of congested streams when average speed of cars is dramatically reduced compared to the maximum allowable speed.

In order to avoid an undesirable mixing between different traffic constellations we apply (similarly to the procedure presented for random matrices in the subsection 1.2) the so-called *traffic unfolding*. Here, such a procedure is composed of two components: a *re-scaling* and *segmentation* by a flux-density window. To be specific, for a certain window-size $(\Delta_J, \Delta_\varrho)$ we introduce a flux-density window $W(\varrho, J) := [\varrho, \varrho + \Delta_\varrho] \times [J, J + \Delta_J]$ that represents a small rectangular sub-region inside the flux-density map. Then the re-scaling is understood as a transformation of the set T_j into the set $\{\tau_{j1}, \tau_{j2}, \dots, \tau_{jm}\}$, where $\tau_{jk} = t_{jk} / \sum_{k=1}^m t_{jk}$. Furthermore, the segmentation

$$I(\varrho, J) = \{j : J_j \in [J, J + \Delta_J] \wedge \varrho_j \in [\varrho, \varrho + \Delta_\varrho]\} \quad (19)$$

is understood as a procedure selecting sub-samples associated with a chosen flux-density window $W(\varrho, J)$. It means that the final sample of unfolded clearances looks like

$$\mathcal{T}(\varrho, J) = \{\tau_{jk} : j \in I(\varrho, J) \wedge k = 1, 2, \dots, m\}. \quad (20)$$

For completeness, we add that in this study there is consistently considered $m = 50$, $\Delta_J = 400 \text{ veh/h}$, and $\Delta_\varrho = 5 \text{ veh/km}$. The total number of analyzed clearances is circa 45 millions (7 record-files gauged on 7 different detectors between 2013-09-16 and 2014-05-29). Among 45 millions of measured clearances 14 millions have been classified (with help of the K-means clustering algorithm) as congested (or transitional).

3.3. Changes in headway statistics induced by varying fluxes and densities

In this subsection we aim to analyze a clearance distribution extracted from unfolded data-sets (20) associated to a fixed flux-density window $W(\varrho, J)$. Withal we try to test a hypothesis that clearance distributions detected in regions of congested traffic belong to the two-parametric family of distributions

$$\wp(\tau) = A\tau^\alpha e^{-\beta/\tau - D\tau}. \quad (21)$$

If proving that hypothesis we find a direct link between inner structures of vehicular ensembles and spectra of damped random matrices.

The testing mechanism is as follows. Varying a flux-density window we estimate a optimal parameters $\alpha, \beta \geq 0$ for which the statistical distance

$$\chi(\alpha, \beta) := d[\wp, q] = \sqrt{\int_0^{+\infty} \frac{\pi\tau}{2} e^{-\frac{\pi\tau^2}{4}} |\wp(\tau; (\alpha, \beta)) - q(\tau)|^2 d\tau} \quad (22)$$

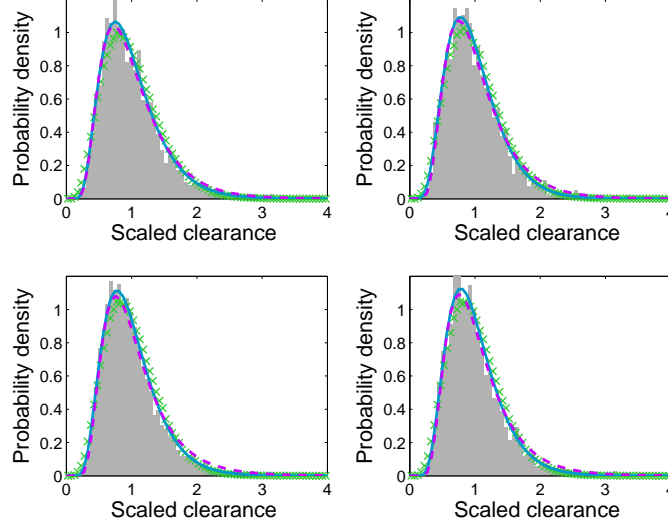


Figure 5. Time clearance distributions in vehicular flows (analyzed for fast lanes of expressways). Histograms stand for empirical distributions analyzed inside the following flux-density windows: a northwest sub-figure: $W(48, 2300)$, northeast sub-figure: $W(54, 1900)$, southwest sub-figure: $W(58, 1300)$, and southeast sub-figure: $W(62, 1300)$. The continuous/dashed curves visualize functions (21) specified for parameters estimated by MDE/MLE, respectively. The crosses correspond to a nearest distribution of eigenvalue-gaps, i.e. to a level spacing distribution whose statistical distance to traffic histogram is minimum.

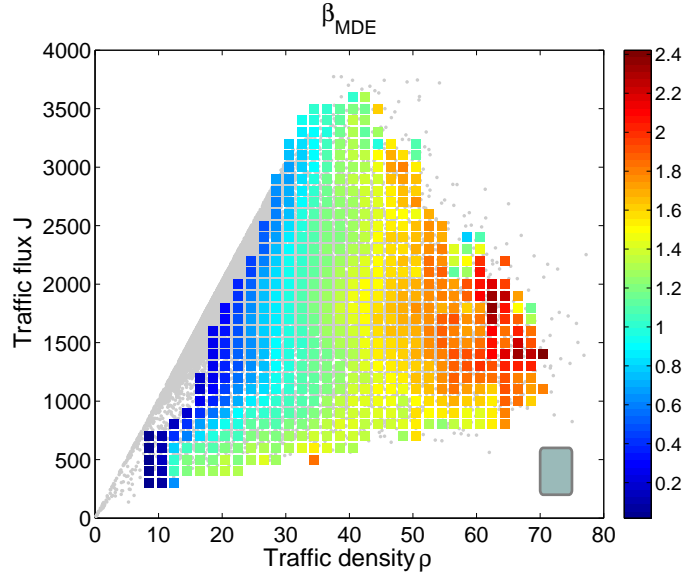


Figure 6. Color representation of values β estimated by minimum distance estimation (23). We plot values $\hat{\beta}_{\text{MDE}}$ in various flux-density windows $W(\varrho, J)$. Squares symbolize locations of window-centroids $(\varrho + \Delta_\varrho/2, J + \Delta_J/2)$. The rectangle at the bottom right schematically visualizes the window-size used. Gray dots correspond to locations (ϱ_k, J_k) of individual data sub-samples before a separation procedure. Unstained areas indicate either a lack of data or belonging to a free-traffic regime. The plotted values are valid for a fast lane.

between a function (21) and clearance histogram $q(\tau)$ is minimum. Practically, after eliminating outliers $\tau_{jk} > 6$ and after a repeated re-scaling (so that the average clearance is equal to one again) we solve a minimization problem

$$(\hat{\alpha}, \hat{\beta})_{\text{MDE}} = \operatorname{argmin}_{\alpha, \beta} \int_0^6 \tau e^{-\frac{\pi \tau^2}{4}} |\wp(\tau; (\alpha, \beta)) - q(\tau)|^2 d\tau, \quad (23)$$

where the histogram $q(\tau)$ is enumerated for the data set $\tilde{T}(\varrho, J) = \{\tilde{\tau}_i : i = 1, 2, \dots, M\}$, for which $\langle \tilde{T}(\varrho, J) \rangle = 1$. Analogously, the maximum likelihood estimations were carried out. For these aims we have defined an estimator

$$\mathcal{L}(\alpha, \beta) = M \ln(A) + \alpha \sum_{i=1}^M \ln(\tilde{\tau}_i) - \beta \sum_{i=1}^M \frac{1}{\tilde{\tau}_i} - D \sum_{i=1}^M \tilde{\tau}_i. \quad (24)$$

By a maximization

$$(\hat{\alpha}, \hat{\beta})_{\text{MLE}} = \operatorname{argmax}_{\alpha, \beta} \mathcal{L}(\alpha, \beta) \quad (25)$$

we obtain alternative values of estimated parameters.

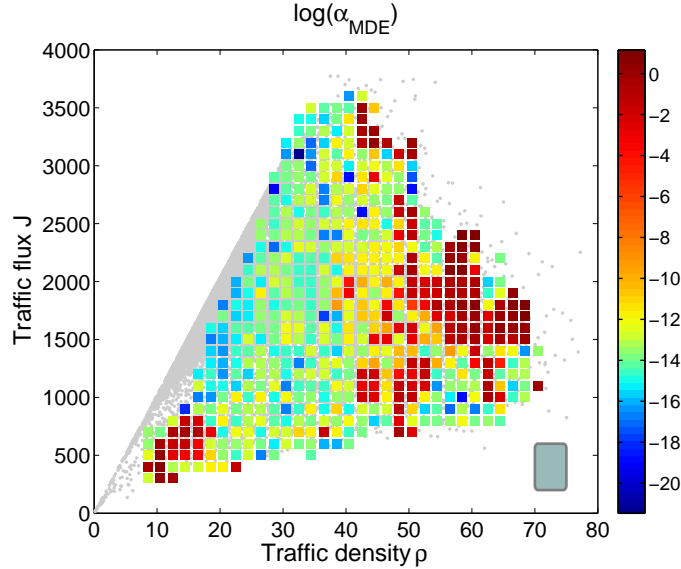


Figure 7. Color representation of values α estimated by minimum distance estimation (23). We plot values $\log(\hat{\alpha}_{\text{MDE}})$ in various flux-density windows $W(\varrho, J)$. Squares symbolize locations of window-centroids $(\varrho + \Delta_\varrho/2, J + \Delta_J/2)$. The plotted values are valid for a fast lane.

In several figures we now depict selected results of our investigations. Method MDE provides estimations visualized in figures 6 and 7. Here one can recognize a behavior so typical for agent systems with intensive interactions among agents. Indeed, with increasing density a level of vehicular coordinations is growing up. This effect is clearly visible in figure 6 where estimated parameters $\hat{\beta}$ are continuously changing from negligible values (for regimes with small densities) to extremely high values detected for a strongly saturated traffic. Such a behavior fully corresponds with results published in the previous researches [8, 6, 9, 16]. Since the parameter β corresponds (in the thermodynamic model interpretation) to the inverse temperature ν discussed in the section 2 we can now provide a deeper insight into changes of a vehicular microstructure. Contrary to the previous investigations, in this paper we present (for the first time, as we believe) a more-detailed study of time-clearance distribution that is performed not in density windows only (as for example in [16, 8]) but in flux-density windows. The presented analysis shows that the principal quantity determining a shape of headway distribution is the traffic density. Deviations caused by changes of a flux J (if density ϱ is fixed) are almost imperceptible. Moreover, estimated values of α are low which confirms the results published formerly. If analyzing the minimum distance $\chi(\hat{\alpha}, \hat{\beta})$ we can see (figure 8) that the tested hypothesis is confirmed predominantly in the region of pure congested flows. A red triangle (boundary of a color area in figure 8) characterizing windows $W(\varrho, J)$ with the

highest deviations can be explained as follows. The east and south sides of the triangle indicate regions where the number of data is not sufficient (and histograms are not therefore smooth enough), whereas the west side represents windows where congested traffic is mixed with free-flow traffic.

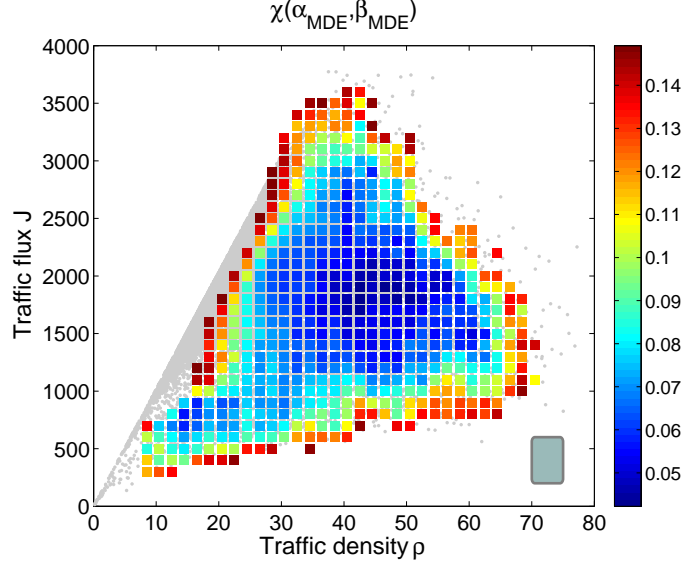


Figure 8. Color representation of a minimum statistical distance $\chi(\alpha, \beta)$ obtained by the MDE-method. We plot values $\chi(\hat{\alpha}_{\text{MDE}}, \hat{\beta}_{\text{MDE}})$ in various flux-density windows $W(\varrho, J)$. Squares symbolize locations of window-centroids $(\varrho + \Delta_\varrho/2, J + \Delta_J/2)$. The plotted values are valid for a fast lane.

For completeness, we note that the results pictured in figures 5, 6, 7, and 8 has been extracted from fast-lane data where car-car interactions are stronger (due to high speeds). Alternative results (valid for main-lane data) are visualized in figures 11, 12, and 13 in the Appendix 5.3. Comparing results obtained for fast and main lanes one can detect significant differences that are particularly noticeable in figures 6 and 11. Here we detect different ranges of parameter β and also different maxima for flux and density. All these disparities are caused predominantly by lower speeds of cars moving in a main lane.

Optimal parameters $\hat{\alpha}, \hat{\beta}$ have been estimated by the method MLE as well. However, their values are (similarly to the results described in the subsection 1.3) extremely close to $\hat{\alpha}_{\text{MDE}}, \hat{\beta}_{\text{MDE}}$ and therefore we will not depict them in figures.

3.4. Damped matrices vs. freeway traffic – direct link

Finally, we analyze a direct link between the random matrices investigated and traffic-flows monitored on expressways. For these intentions we minimize (analogously to the previous approaches) the statistical distance χ calculated for a histogram of vehicular clearances and for a histogram of level spacings in DUE-ensembles. Thus, we search an optimal value of the damped parameter g for which that distance is minimum. Comparison of the above-referred histograms is illustratively visualized in the figure 5 (see histograms and crosses). Optimal values of g are plotted in figure 9. Associated minimum distance are plotted in figure 10. As understandable, the values visualized in figure 10 are higher than those enumerated as a distance between real-road distributions and the functional family (21). The main reason is that the former represents one-parameter fitting, whereas the latter represents an estimation that is two-parametric.

4. Highlights, discussion and conclusion remarks

This article has two scientific layers: mathematical (dealing with new types of random matrices) and physical (investigating properties of inner structure of vehicular streams). Firstly, this paper builds on the pioneering work [13] dealing with new classes of random matrices. It brings a substantially deeper view into a spectral structure of damped matrices than the article [13] itself. Specifically, using theoretical

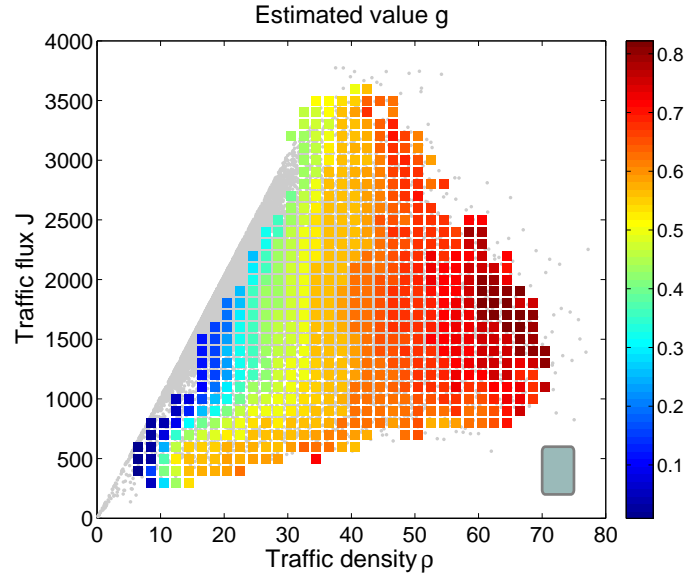


Figure 9. Color representation of the estimated value g obtained by the MDE-method. We plot values $\chi(\hat{\alpha}_{\text{MDE}}, \hat{\beta}_{\text{MDE}})$ in various flux-density windows $W(\varrho, J)$. Squares symbolize locations of window-centroids $(\varrho + \Delta_\varrho/2, J + \Delta_J/2)$. The plotted values are valid for a fast lane.

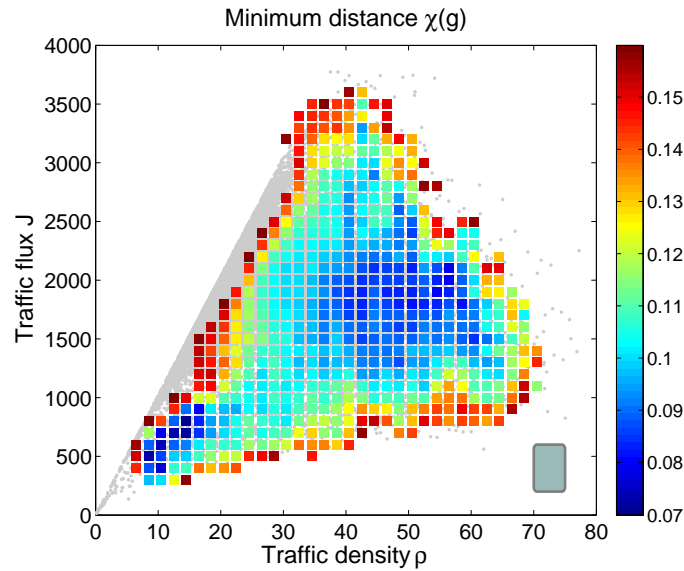


Figure 10. Color representation of a minimum statistical distance $\chi(\alpha, \beta)$ obtained by the MDE-method. We plot values $\chi(\hat{\alpha}_{\text{MDE}}, \hat{\beta}_{\text{MDE}})$ in various flux-density windows $W(\varrho, J)$. Squares symbolize locations of window-centroids $(\varrho + \Delta_\varrho/2, J + \Delta_J/2)$. The plotted values are valid for a fast lane.

approaches we predict an analytical formula for level density of DUE–matrices, describe particularities of the unfolding-procedure, and estimate specific values of parameters used in a predictive formula for the LS-distribution. Parameters in a two-parametric family (13) are estimated with help of the minimum distance estimator and the standard Kolmogorov estimator.

Secondly, we show that the same mathematical formula may be used as a theoretically-substantiated prognosis for a time-clearance distribution describing statistics of time-gaps among consecutive vehicles in real-road traffic. Indeed, both the estimation-procedures used (minimum distance estimation and maximum likelihood estimation) prove that the function (21) represents a suitable candidate for a time-clearance distribution. In this work, for the first time, the time-clearance distribution is monitored in various flux-density sub-windows inside the entire flux-density plane. Thus, the output represents a map of estimated values α and β depending on traffic density and flux. Since the parameter β admits a thermodynamic interpretation (as is discussed in [7, 8]) the changes of the socio-physical coefficient $\beta = \beta(\varrho, J)$ reflect a level of mental pressure which the drivers are under during a given traffic constellation. As is evident in figure 6 for fast lanes and 11 for main lanes, a driver’s mental strain depends predominantly on actual density and only marginally on actual flow. Among other things, we can clearly detect here a process of traffic saturation accompanied by corresponding changes in a traffic-microstructure. More specifically, with increasing density the coefficient β is growing, which is reflected in more systematic configurations of vehicles (as a consequence of the fact that the variance of a clearance distribution drops). Another benefit for physics of traffic is a relatively small statistical distance (if compared with previously published analyzes) between real-road histograms and a theoretical prediction. It confirms a quality of the approach used.

Finally, it has been demonstrated that the new type of random matrix ensembles (g –parameterized class DUE $_g$) may be used to describe statistical distributions of inter-vehicle gaps. To be specific, similarly to the Ref. [2, 3, 4] (where it is proven that time-gap distribution monitored for Cuernavaca buses is the same as the level spacing distribution of GUE–matrices) we now find that a similar correspondence exists between cars on expressways and the ensemble DUE. For a chosen traffic state (described by flux and density) there exists an optimal value of the parameter g which the LS-distribution corresponds for to time-clearance distribution measured on a expressway. Figure (9) shows that the force-parameter g is increasing with traffic density, which supports an intuitive notion on changes of interaction forces in thickening flows.

To be sure that the above-mentioned similarities between traffic and Random Matrix Theory are more substantial (than presented in our article) one should show that the detected correspondence remains valid also if investigating second order characteristics (statistical rigidity or number variance). This will be covered in continuing researches. Preliminary numerical analysis, however, shows that a link between vehicular samples and eigenvalues of random matrices is definitely not accidental.

Acknowledgments

The authors would like to thank The Road and Motorway Directorate of the Czech Republic (Ředitelství silnic a dálnic ČR) for providing traffic data analyzed in this paper.

Research presented in this work was supported by the grant 15-15049S provided by Czech Science Foundation (GA ČR). Partial support was provided by Czech Technical University in Prague within the internal project SGS15/214/OHK4/3T/14.

5. Appendix

5.1. Convergence of the distribution (2)

Let \mathcal{D} be the space of all smooth functions with a bounded support (according to [20]) and \mathcal{D}' be the class of generalized functions (linear and continuous functionals over a space \mathcal{D} .) Then for an arbitrary $\varphi(x) \in \mathcal{D}$:

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0+} (q_{\vartheta, \varepsilon}(x); \varphi(x)) &= \lim_{\varepsilon \rightarrow 0+} \int_{\mathbf{R}} \frac{\zeta(\vartheta)}{\varepsilon} \Theta(\varepsilon - |x|) \exp\left[-\frac{\vartheta^2 \varepsilon^2}{\varepsilon^2 - x^2}\right] \varphi(x) dx = \\ \zeta(\vartheta) \lim_{\varepsilon \rightarrow 0+} \int_{\mathbf{R}} \Theta(1 - |y|) \exp\left[-\frac{\vartheta^2}{1 - y^2}\right] \varphi(\varepsilon y) dy &= \\ \zeta(\vartheta) \varphi(0) \int_{-1}^1 \exp\left[-\frac{\vartheta^2}{1 - y^2}\right] dy &= \varphi(0) = (\delta(x); \varphi(x)). \end{aligned} \quad (26)$$

Therefore $\lim_{\varepsilon \rightarrow 0+} q_{\vartheta, \varepsilon}(x) \stackrel{\mathcal{D}'}{=} \delta(x)$. Analogously $\lim_{\varkappa \rightarrow 0+} \frac{1}{\sqrt{2\pi\varkappa}} e^{-\frac{x^2}{2\varkappa^2}} \stackrel{\mathcal{D}'}{=} \delta(x)$, for

$$\begin{aligned} \lim_{\varkappa \rightarrow 0+} \left(\frac{1}{\sqrt{2\pi\varkappa}} e^{-\frac{x^2}{2\varkappa^2}}; \varphi(x) \right) &= \lim_{\varkappa \rightarrow 0+} \int_{\mathbf{R}} \frac{1}{\sqrt{2\pi\varkappa}} e^{-\frac{x^2}{2\varkappa^2}} \varphi(x) \, dx = \\ \lim_{\varkappa \rightarrow 0+} \int_{\mathbf{R}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \varphi(\varkappa y) \, dy &= \varphi(0) = (\delta(x); \varphi(x)). \end{aligned} \quad (27)$$

For completeness, we add that $\Theta(x)$ stands for the Heaviside step function.

5.2. Normalization and scaling in the distribution (13)

Denoting $J(x, \varkappa) = \int_0^\infty e^{-\frac{x^2}{4t}} e^{-\varkappa^2 t} t^\alpha \, dt$ the normalization and scaling conditions for the density (13) transform into the equation

$$J(x, \varkappa) = -\frac{1}{2\varkappa} \frac{\partial J}{\partial \varkappa}. \quad (28)$$

We add that the substitutions $D = \varkappa^2$, $\beta = x^2/4$ have been applied for convenience. Using the identity $J(x, \varkappa) = \frac{1}{2^\alpha} \left(\frac{x}{\varkappa}\right)^{1+\alpha} \mathcal{K}_{\alpha+1}(x\varkappa)$ one finds

$$(2\varkappa^2 - 1 - \alpha) \mathcal{K}_{\alpha+1}(x\varkappa) = -x\varkappa \mathcal{K}'_{\alpha+1}(x\varkappa). \quad (29)$$

By means of the asymptotical expression

$$\mathcal{K}_m(z) z^m \stackrel{z \rightarrow 0+}{\approx} (2m-2)!! \left(1 + \frac{2z}{2m-1}\right)^{m-\frac{1}{2}} e^{-z}, \quad (2m \geq 1) \quad (30)$$

we obtain the scaling equation for small x :

$$2x\varkappa^3 + \varkappa^2(1 + 2\alpha - x^2) - 2x\varkappa(\alpha + 1) - (\alpha + 1)(2\alpha + 1) = 0. \quad (31)$$

Near the infinity this equation can be simplified to the form

$$4\varkappa^2 - 2x\varkappa - (2\alpha + 3) = 0, \quad (32)$$

which leads to a detection of the linear trend in the scaling constant

$$D(\beta) \approx \beta + \frac{3}{2} + \alpha \quad (\beta \rightarrow \infty). \quad (33)$$

On contrary, with help of the approximating expression

$$\mathcal{K}_m(z) \stackrel{z \rightarrow \infty}{\approx} \sqrt{\frac{\pi}{2}} e^{-z} \left(\frac{1}{\sqrt{z}} + \frac{4m^2 - 1}{8z^{3/2}} \right) \quad (34)$$

we obtain the scaling condition for large x :

$$16x\varkappa^3 + 2\varkappa^2(3 + 8\alpha + 4\alpha^2 - 4x^2) - x\varkappa(15 + 16\alpha + 4\alpha^2) - (4\alpha^3 + 12\alpha^2 + 11\alpha + 3) = 0. \quad (35)$$

Surprisingly, this formula leads (after a simplification) to the scaling formula (32) as well, and therefore to the same linear asymptotics (33).

Above that, since the variant $\beta = 0$ in (13) represents the well-known gamma distribution it is easy to prove that $\lim_{\beta \rightarrow 0+} D(\beta) = \alpha + 1$ and therefore the final scaling formula (phenomenologically corrected) reads (14). Finally, for fixed values α , β , and D we have (15).

5.3. Minimum distance estimations for main lanes

In this part of Appendix we picture results of minimum distance estimations applied to traffic data extracted from main lanes of the Prague Ring.

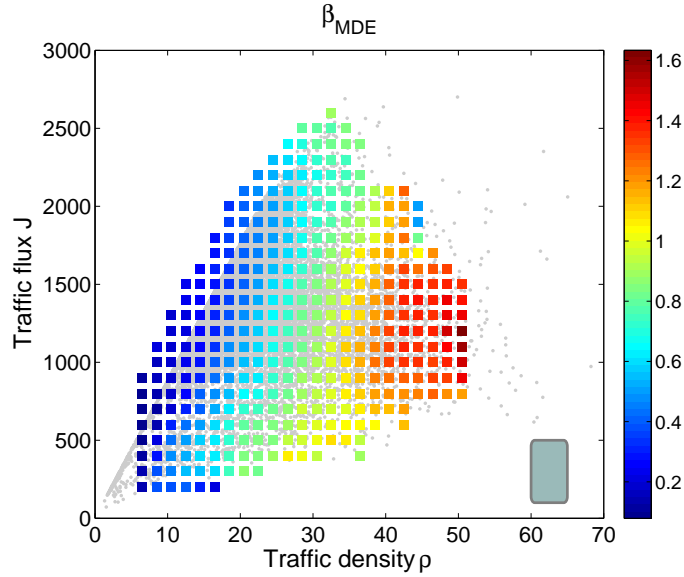


Figure 11. Color representation of values β estimated by minimum distance estimation (23). We plot values $\hat{\beta}_{\text{MDE}}$ in various flux-density windows $W(\varrho, J)$. Squares symbolize locations of window-centroids $(\varrho + \Delta_{\varrho}/2, J + \Delta_J/2)$. The rectangle at the bottom right schematically visualizes the window-size used. Gray dots correspond to locations (ϱ_k, J_k) of individual data sub-samples before a separation procedure. Unstained areas indicate either a lack of data or belonging to a free-traffic regime. The plotted values are valid for a main lane.

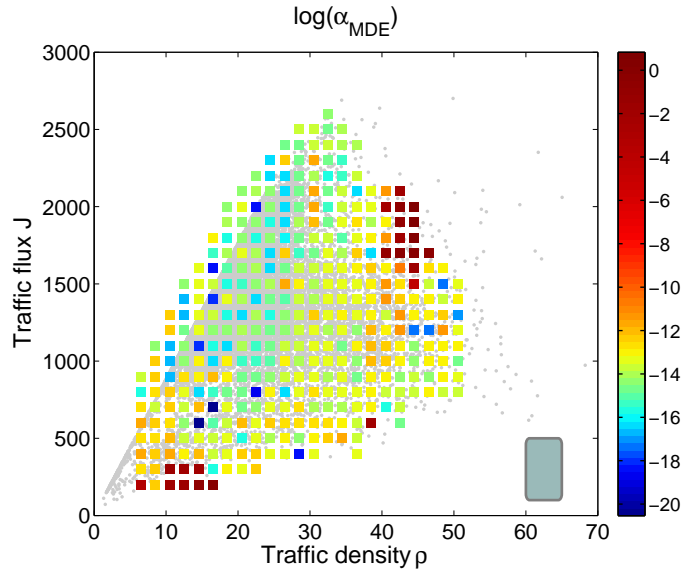


Figure 12. Color representation of values α estimated by minimum distance estimation (23). We plot values $\log(\hat{\alpha}_{\text{MDE}})$ in various flux-density windows $W(\varrho, J)$. Squares symbolize locations of window-centroids $(\varrho + \Delta_{\varrho}/2, J + \Delta_J/2)$. The plotted values are valid for a main lane.

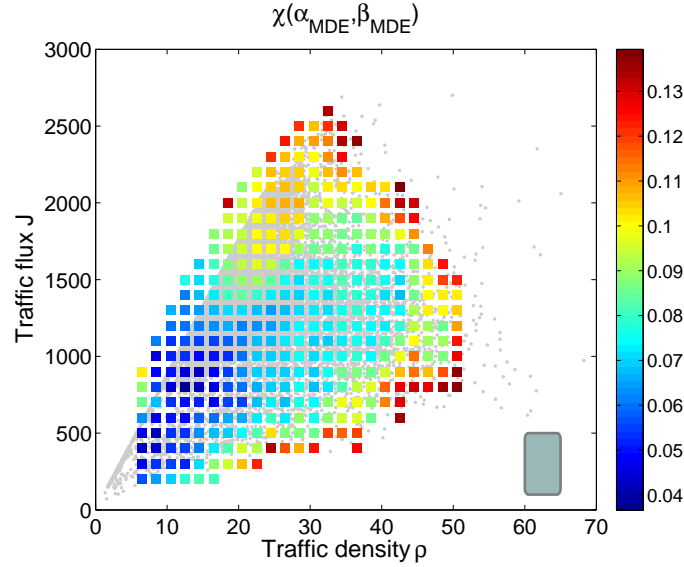


Figure 13. Color representation of a minimum statistical distance $\chi(\alpha, \beta)$ obtained by the MDE-method. We plot values $\chi(\hat{\alpha}_{\text{MDE}}, \hat{\beta}_{\text{MDE}})$ in various flux-density windows $W(\varrho, J)$. Squares symbolize locations of window-centroids $(\varrho + \Delta_\varrho/2, J + \Delta_J/2)$. The plotted values are valid for a main lane.

References

- [1] Tao T 2014 *Komplexität und Universalität* E-enterprise Lemgo
- [2] Krbálek M and Šeba P 2000 J. Phys. A: Math. Gen. **33** L229
- [3] Krbálek M and Šeba P 2003 J. Phys. A: Math. Gen. **36** L7
- [4] Baik J, Borodin A, Deift P and Suidan T 2006 J. Phys. A: Math. Gen. **39** (2006) 8965
- [5] Krbálek M and Helbing D 2004 Physica A **333** 370
- [6] Krbálek M 2008 J. Phys. A: Math. Theor. **41** 205004
- [7] Krbálek M and Šeba P 2009 J. Phys. A: Math. Theor. **42** 345001
- [8] Krbálek M 2013 J. Phys. A: Math. Theor. **46** 445101
- [9] Krbálek M and Šleis J 2015 J. Phys. A: Math. Theor. **48** 015101
- [10] May A D 1990 *Traffic Flow Fundamentals* Englewood Cliffs, NJ: Prentice-Hall
- [11] Jin X, Zhang Y, Wang F, Li L, Yao D, Su Y and Wei Z 2009 Transportation Research Part C: Emerging Technologies **17**(3) 318
- [12] Li L, Wang F, Jiang R, Hu J and Ji Y 2010 Chinese Phys. B **19** 020513
- [13] Bogomolny E, Giraud O and Schmit C 2011 Nonlinearity **24** 3179
- [14] Mehta M L 2004 *Random matrices (Third Edition)* New York: Academic Press
- [15] Dodge Y 2003 *The Oxford dictionary of statistical terms* New York: Oxford University Press
- [16] Krbálek M 2007 J. Phys. A: Math. Theor. **40** 5813
- [17] Hartigan A 1975 *Clustering algorithms* John Wiley & Sons
- [18] Treiber M and Kesting A 2013 *Traffic Flow Dynamics* Springer
- [19] Kerner B S 2004 *The Physics of Traffic* Berlin, New York: Springer Verlag
- [20] Vladimirov V. S. 1971 *Equations of Mathematical Physics* New York: Dekker